

***TBX6*-associated congenital scoliosis (TACS) as a clinically distinguishable subtype of congenital scoliosis: further evidence supporting the compound inheritance and *TBX6* gene dosage model**

**Supplementary Materials**

Materials and Methods

Fig. S1. Demographics of Discovery Cohort 1.

Fig. S2. High-density CGH Microarray Identified Seven 16p11.2 Deletions in the Newly Recruited Cases from Cohort 1.

Fig. S3. Droplet Digital PCR (ddPCR) Analysis Identified Two 16p11.2 Deletions in Cohort 1.

Fig. S4. Sanger Sequencing Validated One Nonsense Mutation in Cohort 1, One Splice Site Mutation in Cohort 2 and One Frameshift Mutation in Cohort 3.

Fig. S5. 16p11.2 Deletions Identified in Five Japanese Sporadic CS Cases from Cohort 2.

Fig. S6. Distribution of TACS in CS Cohort 1.

Fig. S7. Mouse *Tbx6* LoF and Mild-hypomorphic Mutation Generated by CRISPR/Cas9.

Fig. S8. Spinal Phenotype of *Tbx6* Gene-Edited Mice.

Fig. S9. Vertebral Malformation in a TACS Patient Treated by Surgical Correction at an Early Age.

Table S1. The Methods of Genetic Analyses in the Three Cohorts.

Table S2. Genotype and Demographic Information of all TACS Patients in Cohort 1 and Cohort 2.

Table S3. Mouse Phenotype with Different *Tbx6* Genotypes.

Table S4. Parameters in the Multivariable TACS Model.

Table S5. Using Points $\geq 3$  as the Cutoff Value in the Discovery Cohort 1.

Table S6. Using Points $\geq 3$  as the Cutoff Value in the Validation Cohort 2.

## **Materials and Methods**

### *DNA Sample Preparation*

In Cohort 1, five ml of peripheral blood was collected from each participant. Genomic DNA was extracted with the QIAamp DNA Blood Mini Kit (QIAGEN, Germany) according to manufacturer's instructions. DNA quality was assessed by agarose gel electrophoresis. DNA samples were quantified by Qubit BR measurement (Invitrogen, USA).

In Cohort 2, genomic DNA was extracted from peripheral blood leukocytes using standard methods or from saliva using the Origen DNA collection kit (DNA Genotek, Canada).

### *Comparative Genomic Hybridization Microarrays*

The genome-wide copy number variant analyses were conducted using high-density ( $1 \times 1$  M) Agilent oligonucleotide-based comparative genomic hybridization (CGH) microarrays. Customized CGH microarrays were used to confirm some of the deletion candidates indicated by qPCR in Cohort 1 with a deletion at 16p11.2. The experimental details were previously described.<sup>1</sup>

The genomic DNA extracted from peripheral blood leukocytes of each subject and the gender-matched reference DNA (Promega, USA) were respectively fragmented using AluI and RsaI enzyme digestion. DNA labeling was conducted using Agilent SureTag DNA Labeling Kit. Different fluorescence dyes were used for DNA labeling of each subject (Cy5-dUTP) and the reference DNA (Cy3-dUTP). Each labeled subject DNA was hybridized together with the labeled

reference DNA onto Agilent human CGH microarrays (including the formats of 1×1M, 1×244k, and 4×180k).

The Agilent 1×1M and 1×244k CGH microarrays are commercially available, which have approximately one million and 244,000 oligonucleotide probes, respectively. These two formats were used for genome-wide CNV analysis in this study. In addition, a custom 4×180k CGH microarray with approximately 180,000 oligonucleotide probes was used for verification of the 16p11.2 deletions suggested by quantitative PCR. This custom microarray has 509 probes in the 16p11.2 deletion region and achieves a much higher probe density in the target region than the Agilent 1×1M and 1×244k CGH microarrays. In addition to 16p11.2, our custom CGH microarray also covers 360 additional neurological genes/loci that are unrelated to this study.

DNA processing, microarray handling, and data analyses were conducted by following the Agilent oligonucleotide CGH protocol (version 6.0). The genome-wide CNV analysis was also conducted in control subjects of Han Chinese using Agilent 1×1M and/or NimbleGen 4.2M (Roche, USA; with approximately 4.2 million probes covering the whole human genome) CGH microarrays.

#### *Quantitative PCR Analysis*

Quantitative polymerase chain reaction (qPCR) analysis was conducted to screen for 16p11.2 deletions in Cohort 1. Two test loci (named PA and PB) in the 16p11.2 deletion region and one reference locus (named P1) outside of the deletion region were used. The qPCR primers are provided in the table below. The experiments were conducted using the SYBR Green Real-

time PCR Master Mix (TOYOBO, Japan) and ABI Prism™ 7900HT Sequence Detection System. Three replicates were conducted for each assay. The average Ct value,  $\Delta\text{Ct}$  (PA-P1) and  $\Delta\text{Ct}$  (PB-P1) were calculated for each sample. The experimental details were previously described.<sup>1</sup> Any 16p11.2 deletion candidates suggested by qPCR were confirmed by CGH microarray or digital droplet PCR (ddPCR).

#### **Primers for Quantitative PCR and Digital Droplet PCR in Cohort 1.**

| <b>Primer</b> | <b>Sequence (5'-3')</b> |
|---------------|-------------------------|
| P1-F          | GGGGAAGGAACTTACATGAC    |
| P1-R          | TCGTGTTTCCCTGTTGTACC    |
| PA-F          | GGTCTAAGCCACACACTAAC    |
| PA-R          | TGAGTTTAGGGACCAATCTA    |
| PB-F          | GCTGCCAGTATGTGACCGAGA   |
| PB-R          | GGGTGGAGGAGAGGATAGGG    |

A TaqMan real-time qPCR method was used to detect the *TBX6* copy number in the congenital scoliosis (CS) patients from Cohort 2 as previously described.<sup>2</sup> The probe, 5'-CTGGGACCAGAGTATT-3', and primers (shown in the table below) were designed against the 3' UTR of *TBX6*. All assays were performed with the TaqMan Universal PCR Master Mix (Applied Biosystems, USA), and the data were analyzed with Copy Caller Software v2.0 (Applied Biosystems, USA).

#### **Primers for TaqMan Real-time qPCR in Cohort 2.**

| <b>Primer</b> | <b>Sequence (5'-3')</b> |
|---------------|-------------------------|
| Forward       | GCAGGCCACACCAGTCTGT     |
| Reverse       | CTGGGACCAGAGTATT        |

### *Digital Droplet PCR (ddPCR)*

For two subjects, namely, XH330 and XH623, ddPCR was conducted to confirm the 16p11.2 deletions. Based on the Bio-Rad Droplet Digital<sup>TM</sup> PCR system, we included the 2× EvaGreen ddPCR Supermix (Bio-Rad, USA) and primers at a final concentration of 0.1 μM in 20-μl PCR reaction mixtures. The ddPCR primers are the same with qPCR primers. Genomic DNA was loaded into all reactions except for assays examining fluorescence intensity as a function of the amount of template DNA. We partitioned each reaction mixture into approximately 20,000 droplets with a droplet generator (Bio-Rad, USA), and the reactions were cycled with the following conditions: 95 °C for 5 min (1 cycle); 95 °C for 30 sec and 52–62 °C for 1 min (40 cycles); 4 °C for 5 min, 90 °C for 5 min (1 cycle), and 4 °C (hold). Cycled droplets were read individually with a droplet reader (Bio-Rad, USA). The copy number was calculated as the ratio of the target (PA and PB) copies to the reference (P1) copies with Bio-Rad QuantaSoft<sup>TM</sup> software.

### *Sanger Sequencing of the TBX6 Gene*

All exons of *TBX6* and the approximately 1-kb upstream region of *TBX6* were investigated. We used the same experimental conditions as described previously to screen the new patients in Cohort 1<sup>1</sup>. The details are provided in the table below.

#### **Primers for Long-range PCR to Amplify the Whole *TBX6* Gene.**

| <b>Primer</b> | <b>Sequence (5'-3')</b>   |
|---------------|---------------------------|
| Forward       | TAGGGAGAGGGCTCTGTTCTCATGG |
| Reverse       | GCGTCCCAGGGAGGCAACCG      |

**PCR Reaction Condition for Long-range PCR.**

| Temperature/ °C | Time   | Cycle |
|-----------------|--------|-------|
| 98              | 1 min  |       |
| 98              | 10 sec |       |
| 60              | 20 sec | 35    |
| 68              | 4 min  |       |
| 68              | 10 min |       |
| 12              | ∞      |       |

**PCR System for the Long-range PCR.**

| Components                                   | Volume/μl |
|--|-----------|
| Water  | 29.5      |
| 10× LA PCR Buffer II (Mg <sup>2+</sup> Plus) | 5         |
| dNTP Mixture (each 2.5 mM)                   | 8         |
| Forward primer (10 μM)                       | 2         |
| Reverse primer (10 μM)                       | 2         |
| DMSO   | 1         |
| Template (50 ng/μl)                          | 2         |
| TaKaRa LA Taq (5U/μl)                        | 0.5       |
| Total  | 50        |

**Sequencing Primers for the Whole *TBX6* Gene.**

| Primer for <i>TBX6</i> fragment | Sequence (5'-3')      |
|---------------------------------|-----------------------|
| Upstream region                 | CTCGAAGGGGTCCGAGAGG   |
| Upstream region and Exon 1      | CTCCTTCCATAGCTCCCGGT  |
| Exon 2                          | GTTGCATACTGATCCCGAAT  |
| Exon 3a                         | CTGCCCCGAAGTAGGTGTATG |
| Exons 3b-5                      | AATGGCTTCCTAACAGATGAC |
| Exons 6-8a                      | GAGCGGGAGGTTTGTGATG   |
| Exon 8b-3'-UTR                  | GGCAGCTGGAAACACAGGT   |
| Exon 9-10-R                     | CTCCCTAGCCCCAACCTCTCC |

All exons and exon-intron junctions of *TBX6* were sequenced for both strands using KOD Fx (TOYOBO, Japan), the primer sets and a 3730xl DNA analyzer (Applied Biosystems, USA) in 121 patients in Cohort 2, with the same experimental conditions as previously described.<sup>2</sup> The *TBX6* variants were also detected by exome sequencing and verified by Sanger sequencing in the remaining 21 patients from Cohort 2 and in BH8084 from Cohort 3. For the remaining nine patients in Cohort 3, all exons of *TBX6* and the approximately 1-kb upstream region of *TBX6* were also investigated by Sanger sequencing using the same method in Cohort 1. According to our previous findings,<sup>1</sup> we analyzed three SNPs at rs2289292, rs3809624, and rs3809627 by Sanger sequencing in patients with 16p11.2 deletions or *TBX6* Loss-of-Function (LoF) mutations.

#### *Haplotyping of Common TBX6 Variants*

For the patients with *TBX6* frameshift or nonsense mutations, haplotyping of the common *TBX6* variants was conducted and verified by Sanger sequencing using the primers shown in the table below. For the new patients in Cohort 1, we used the same experimental conditions as previously described.<sup>1</sup> Specifically, the haplotyping of the common *TBX6* variants was conducted using the ClonExpress One Step Cloning Kit (Vazyme, China). We amplified both the vectors and insert DNA fragments to perform recombination cloning. The pGEM-T vector (Promega, USA) was used as the template for amplifying the vectors. Both the vector and insert DNA fragments were amplified using Q5 polymerase (NEB, UK). The PCR products were purified with a gel exaction kit (GENERAY, China). After ligation, the products were transfected into *E. coli*

competent cells. The bacterial colonies were collected and verified by Sanger DNA sequencing.

#### **Primers for Haplotyping Analysis of the *TBX6* Risk Haplotype in Cohort 1.**

| <b>Primer</b> | <b>Sequence (5'-3')</b>                      |
|---------------|--|
| T7-reverse    | TCGCCCTATAGTGAGTCGTATTACA                    |
| SP6-reverse   | GTATTCTATAGTGTCACCTAAATAG                    |
| CS-F          | GACTCACTATAGGGCGAGGGGAAGGGAGCGGGAGGTTTGTG    |
| CS-R          | GGTGACACTATAGAATACGCGCTGAGCCTGCCGGGAAGTGTAGT |

In Cohort 2, we examined the haplotype using long-range PCR and sequenced the subcloned amplicons.<sup>2</sup> The genomic region of the *TBX6* gene, including the three SNPs, was PCR-amplified using the primers shown in the table below. The PCR amplicons were cloned into the pBluescript SK(-) cloning vector (Agilent Technologies, USA) and sequenced.

#### **Primers for the Haplotyping Analysis of the *TBX6* Risk Haplotype in Cohort 2.**

| <b>Primer</b> | <b>Sequence (5'-3')</b>                 |
|---------------|---|
| Forward       | CCAGTCTCGAGCGCCCCTATAGACCTGTGCGGTTCTGAG |
| Reverse       | CCAGTAAGCTTGAGGGCTCTGTTCTCATGGGCAGCCTTA |

#### *Phenotype Evaluation*

##### *The age of onset*

The age-of-onset is the clinical setting indicating the age at which an individual first came to medical attention, exhibiting or presenting with features such as asymmetric shoulder height, shoulder blade prominence, uneven musculature on one side of the spine, rib prominence, uneven hips, and discrepant arms or leg lengths.

### *Vertebral anomaly.*

The vertebral defects were classified as the defect of formation, defect of segmentation and the mixed defect.<sup>3,4</sup> The abnormal vertebrae were divided into hemivertebrae or hypoplasia, block vertebrae, butterfly vertebrae, non-segmented hemivertebrae, non-segmented butterfly vertebrae and missing vertebrae. Hemivertebra results from a complete defect of the formation with the absence of one pedicle and a region of the vertebral body. Hemivertebra can be further classified on the basis of the presence or absence of fusion to the vertebral bodies above and/or below.<sup>5</sup> A non-segmented hemivertebra is fused to the vertebral body above and/or below, and a segmented hemivertebra is separated from the body above and below by disk space. Hypoplasia, also called wedged vertebra, is resulted from an incomplete defect of formation, which has asymmetry in height, with bilateral pedicles and one side being hypoplastic. A block vertebra results from bilateral and symmetrical bony connections. The abnormal vertebrae were also classified as butterfly vertebrae and non-segmented butterfly vertebrae according to the specific sagittal cleft shape. A missing vertebra was decreasing in the number of vertebrae and only found in T12 region.

### *Rib anomaly.*

Rib defects were classified as structural changes (fused rib, bifid rib, and irregular rib) and number changes. We divided the rib anomalies into simple or complex types following the criteria described by Tsirikos and McMaster.<sup>6</sup> A patient with a simple rib anomaly had only one of the following: a localized fusion of two or three ribs or a small chest wall defect that was due

to a deviation of one or two ribs or an absence of a rib. A patient with complex rib anomalies had multiple extensive rib fusions, usually without a set pattern, combined with an adjacent large chest wall defect that was due to an absence or deviation of ribs.

#### *Intraspinal anomaly.*

Intraspinal anomalies were classified as diastematomyelia, syringomyelia, tethered cord, low conus, arachnoid cyst, Chiari malformation, intraspinal masses and other intraspinal defects.

#### *Cardiac anomaly.*

Cardiac anomalies included mitral valve prolapsed (MVP), atrial septal defect (ASD), ventricular septal defect (VSD), bicuspid aortic valve, patent ductus arteriosus (PDA) and other defects in the heart.

#### *The anomaly of other organs.*

We also identified less number of urogenital, musculoskeletal, auricular and gastrointestinal anomalies in our cohort, which were not shown in this study in detail.

#### *Mouse Model Preparation*

##### *Genome editing in mice using CRISPR/Cas9.*

Zygotes from the FVB mice were edited by CRISPR/Cas9 to generate the mouse model with a *Tbx6* LoF mutation *in trans* with the hypomorphic haplotype on the second allele.<sup>7</sup> The *Tbx6* LoF mutation was prepared by introducing a LoF mutation in exon 2, and the mild-hypomorphic

(mh) allele of *Tbx6* was prepared by editing the T binding site in the *Tbx6* promoter (Figure S7A&B).<sup>7</sup> Cas9 mRNA and gRNA were pooled (10 ng/μl for each RNA) and injected into the zygotes from the FVB mice (200 cells for each edit). The founder (F0) mice were crossed with wild-type FVB mice to obtain offspring (F1). The *Tbx6*<sup>mh/-</sup> mice were obtained by crossing the *Tbx6*<sup>wt/-</sup> mice with the *Tbx6*<sup>wt/mh</sup> mice. Genomic DNA was extracted from the tissues of the toe and used for genotyping by Sanger sequencing. The targets of the guide RNAs (gRNAs) in the mouse genome and the PCR primers used for amplification and sequencing are shown in the table below. The mouse study was carried out in accordance with the recommendation in the Guide for the Care and Use of Laboratory Animals of the USA National Institutes of Health.

#### Target Sequences of Single-guide RNA (sgRNA) of *Tbx6* in the Mouse Genome.

| Location of target region      | Sequences (5'-3')  |
|--------------------------------|--|
| Exon 2                         | <u>GAGGCCCTTCACTCGCT</u> <span style="border: 1px solid black; padding: 0 2px;">TCC</span> <sup>a</sup>    |
| T binding site in the promoter | <u>CCTAGGGATCAATCAGACCG</u> <span style="border: 1px solid black; padding: 0 2px;">AGG</span> <sup>a</sup> |

<sup>a</sup> Sequences of the sgRNA targets are underlined and the protospacer adjacent motif sequences are shown in the boxes.

#### Primers for Mouse *Tbx6* Genotyping.

| Primer                       | Sequences (5'-3')      |
|------------------------------|------------------------|
| <i>Tbx6</i> <sup>mh</sup> -F | CAAGCGTCAGAGAAGACCACCA |
| <i>Tbx6</i> <sup>mh</sup> -R | CCTTACCCAGAGCCAATCCAAC |
| <i>Tbx6</i> -exon 2-F        | ATCCTCATTCTTCCCACATCTC |
| <i>Tbx6</i> -exon 2-R        | ATGCCACCTCTTACAGTTTCT  |

Abbreviation: mh, mild-hypomorphic.

#### *Cell culture and luciferase reporter assay.*

P19CL6 cells were cultured and induced to differentiate into cardiomyocytes as described previously.<sup>8</sup> The cells were harvested every 24 h after treatment with DMSO. We amplified the DNA fragments that included the potential regulatory elements of mouse *Tbx6* to construct reporter plasmids. The fragments obtained from the *Tbx6* wild-type and mild-hypomorphic alleles were separately inserted into the PGL3-Basic vector (Promega, USA), enabling fusion to the reporter gene. The luciferase reporter assay was conducted as described in our previous study.<sup>1</sup>

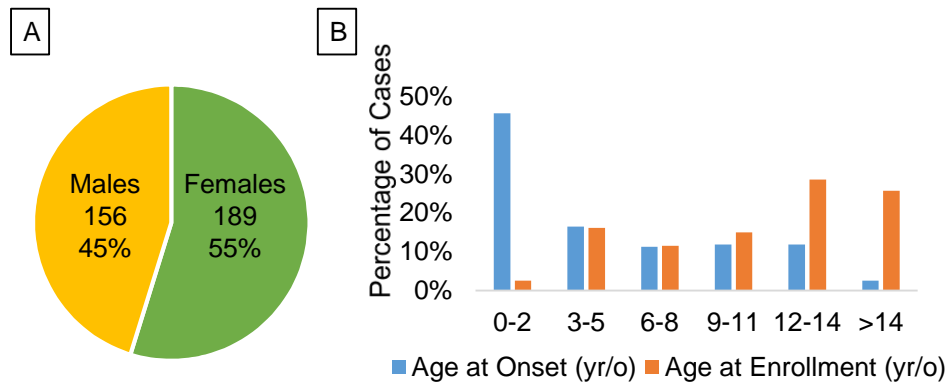
#### *Micro-CT imaging.*

The micro-CT scans of the *Tbx6*<sup>mh/-</sup> mice were conducted when the mice were 35-45 days old to assess the morphology of the vertebrae and ribs. During the scan, each adult mouse was anesthetized by intra-peritoneal injection of 2% pelltobarbitalum natricum. Three-dimensional (3D) datasets were acquired using a Skyscan 1176 high-resolution micro-CT scanner (Bruker, Belgium), and the image projections were reconstructed with the GPURecon Server. The scanning parameters were set as follows: a spatial resolution of 35- $\mu$ m pixels, an X-ray source set at 50 kVp and 497  $\mu$ A, a 0.5-mm Al filter, a 2° rotation step and a 360° rotation around the vertical axis.

#### *Statistical Analyses*

The Mann-Whitney U test was used for statistical analysis of the age at onset and the numbers of malformations among patients with different genotypes. The gender and prevalence

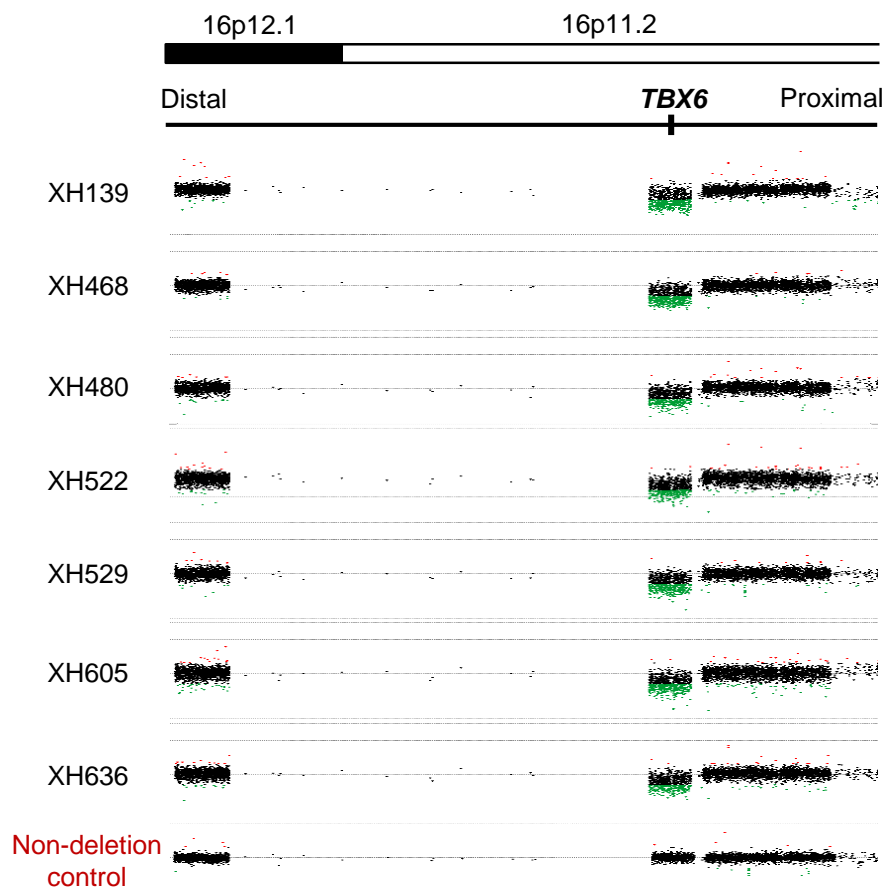
of different types of defects were compared using Pearson's  $\chi^2$  test or Fisher's exact test (two-sided). The odds ratio (OR) with the 95% confidence interval (CI) was used to assess the influence of *Tbx6* LoF mutations on vertebral, rib and other malformations. The unpaired t-test was used for statistical analysis of the qPCR and luciferase reporter assays. Statistical analysis was performed with SPSS version 15.0 (SPSS, USA). Differences with  $P < 0.05$  were considered statistically significant. Additionally, a multivariable model of the *TBX6*-associated CS (TACS) risk score (TACScore) was derived from the Cohort 1 using the model by Sullivan et al <sup>9</sup> to predict the TACS risk for the CS patients using logistic regression, in which points were assigned to each predictive variable with point totals corresponding to the risk estimate. The score of each variable was assigned according to the product of the corresponding  $\beta$  coefficient and the value of the variable. The receiver operating characteristic (ROC) was used to assess the effectiveness of this model in the Chinese derivation cohort and the Japanese replication cohort, in which the sensitivity (SE) was plotted as a function of 1-specificity (1-SP). As one of the main summary statistics of the ROC curve, the Youden index (J) was calculated to define the maximum potential effectiveness of the predictive models, which could be formally defined as  $J = \max_c [SE(c) + SP(c) - 1]$ .<sup>10</sup> Cutoff points for the TACScore were determined with the Youden index and the consensus at a panel meeting.



**Figure S1. Demographics of Discovery Cohort 1.**

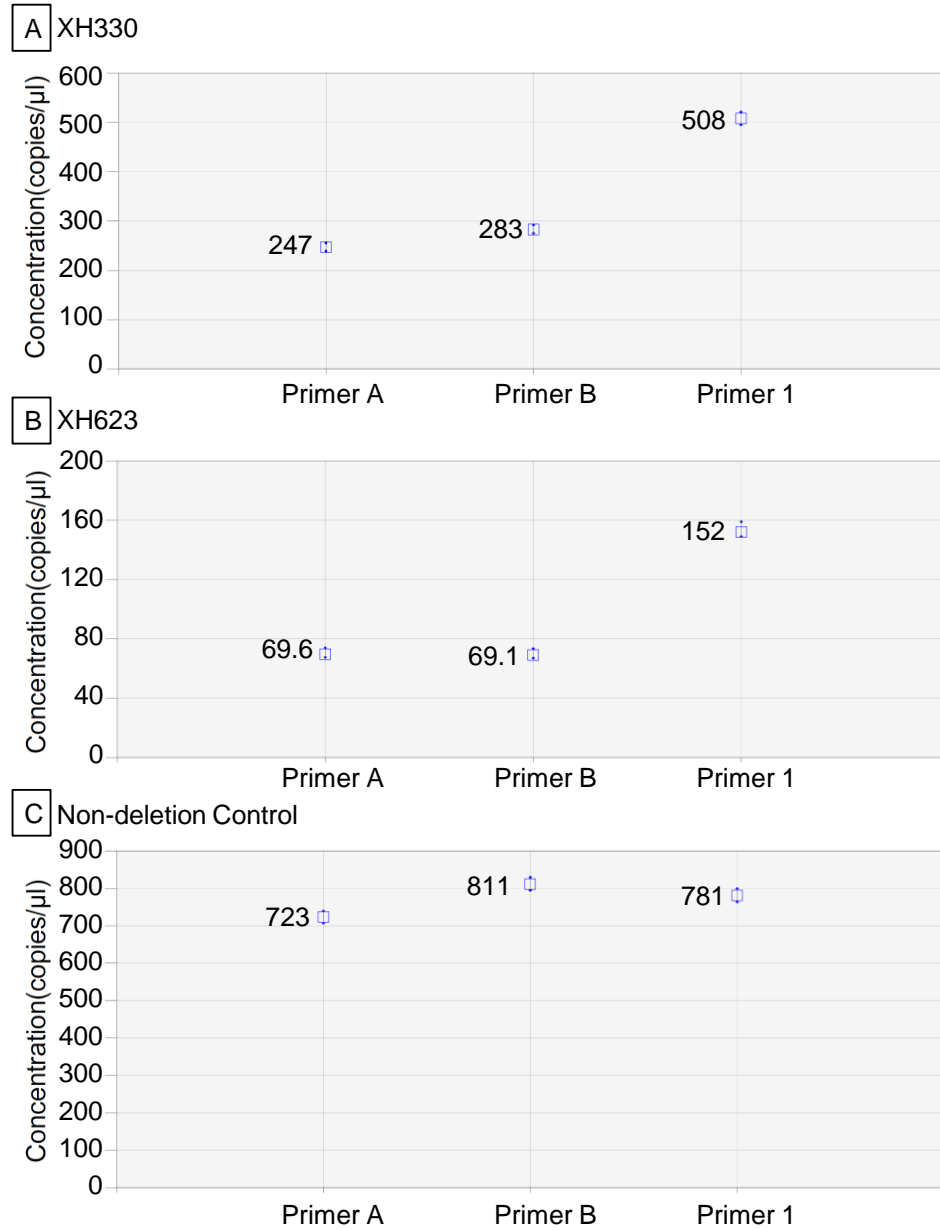
**(A)** Gender distribution. **(B)** The distribution of age at onset and enrollment.

Abbreviation: yr/o, year(s) old.



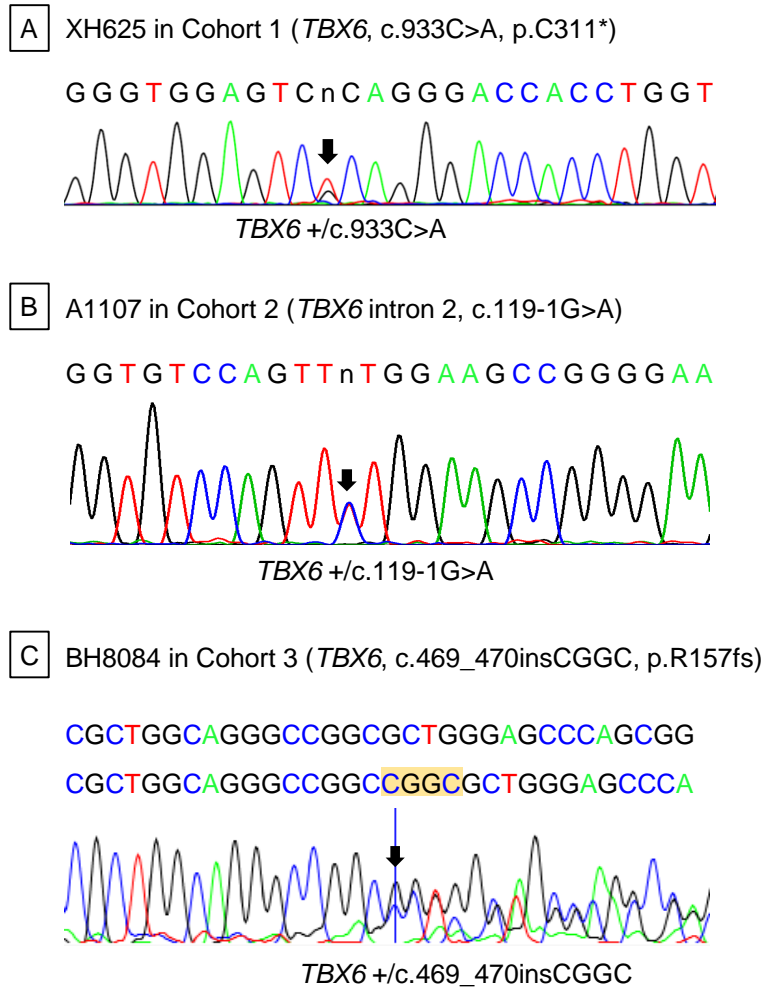
**Figure S2. High-density CGH Microarray Identified Seven 16p11.2 Deletions in the Newly Recruited Cases from Cohort 1.**

These deletions were recurrent. Seven deletions initially suggested by qPCR were confirmed by our custom CGH microarray with 509 oligonucleotide probes located in the deletion region.

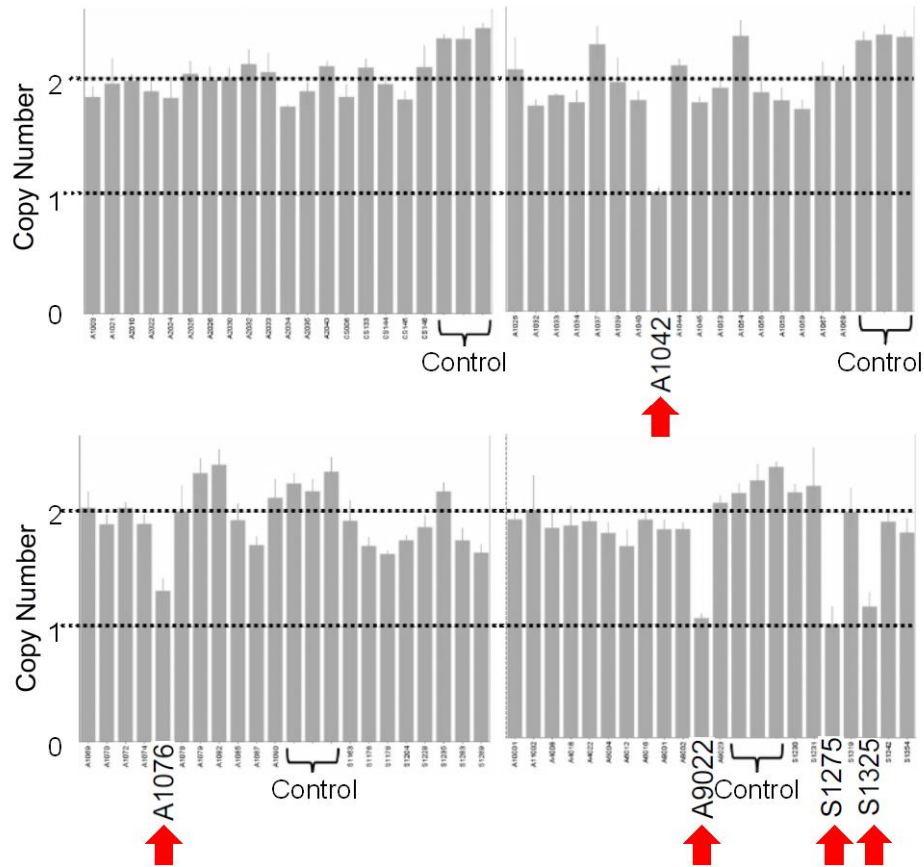


**Figure S3. Droplet Digital PCR (ddPCR) Analysis Identified Two 16p11.2 Deletions in Cohort 1.**

(A-C) Two deletions initially suggested by qPCR were confirmed by ddPCR. The input DNA concentration in copies/μl for the ddPCR experiments were determined with QuantaSoft Analysis Software. The DNA concentrations in two test loci (PA and PB) in the 16p11.2 deletion region were about half of the ones in reference locus (P1) outside of the deletion region in XH330 and XH623.



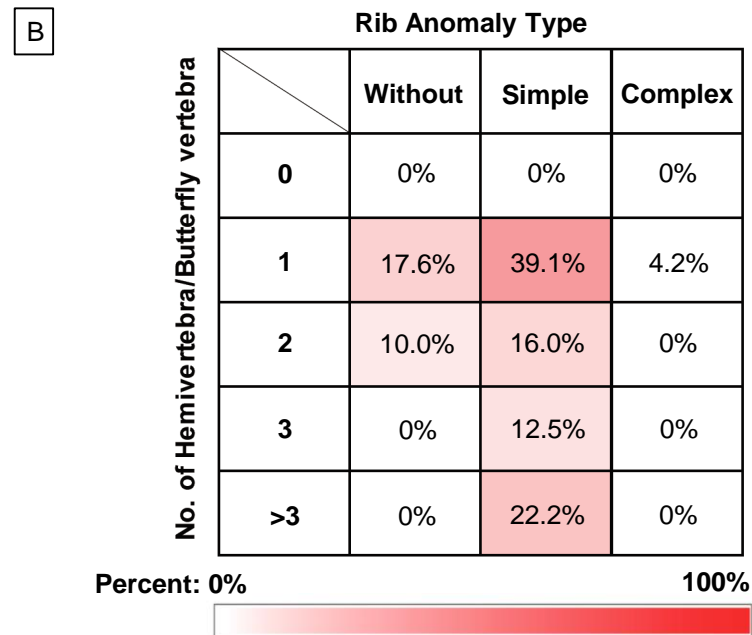
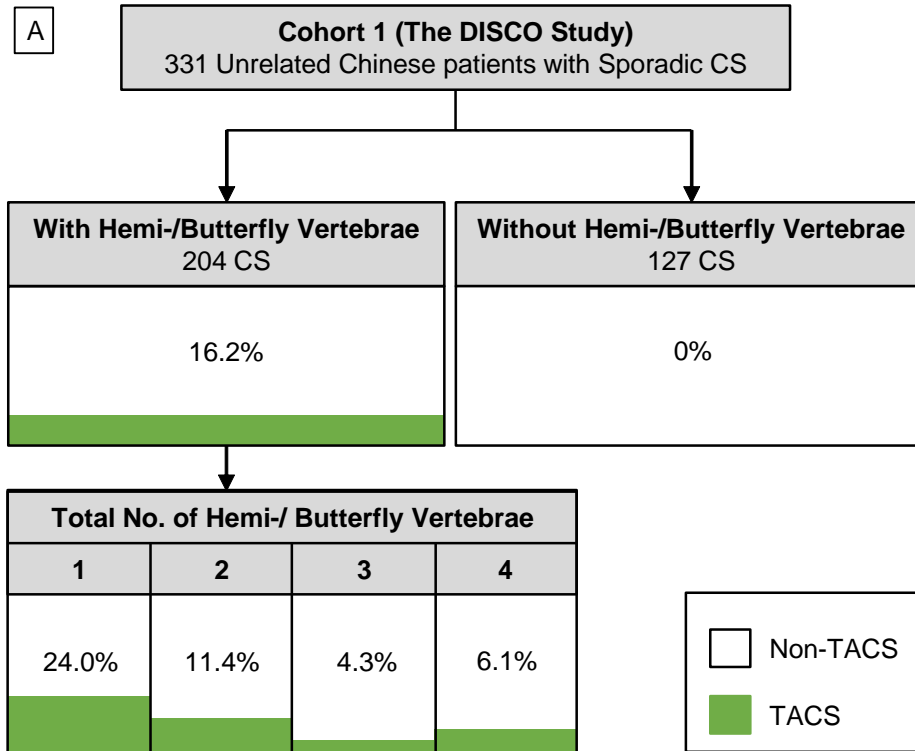
**Figure S4. Sanger Sequencing Validated One Nonsense Mutation in Cohort 1, One Splice Site Mutation in Cohort 2 and One Frameshift Mutation in Cohort 3. (A)** A stop-gain mutation, c.933C>A, was identified in patient XH625 from Cohort 1 by Sanger sequencing. **(B)** A splice site mutation, c.119-1G>A, was identified in patient A1107 from Cohort 2 by Sanger sequencing. **(C)** A frameshift mutation, c.469\_470insCGGC, was identified in patient BH8084 from Cohort 3 by exome sequencing and verified by Sanger sequencing. The black arrows indicate the nucleotide substitution positions of each *TBX6* variant. The ‘+’ indicates the human genome reference.



**Figure S5. 16p11.2 Deletions Identified in Five Japanese Sporadic CS Cases from Cohort**

**2.**

The 16p11.2/*TBX6* deletion was detected by TaqMan real-time qPCR. Five CS patients (A1042, A1076, A9022, S1275, and S1325) were determined to have one copy, and the other patients and normal controls were determined to have two copies. The red arrows indicate the 16p11.2/*TBX6* deletions.

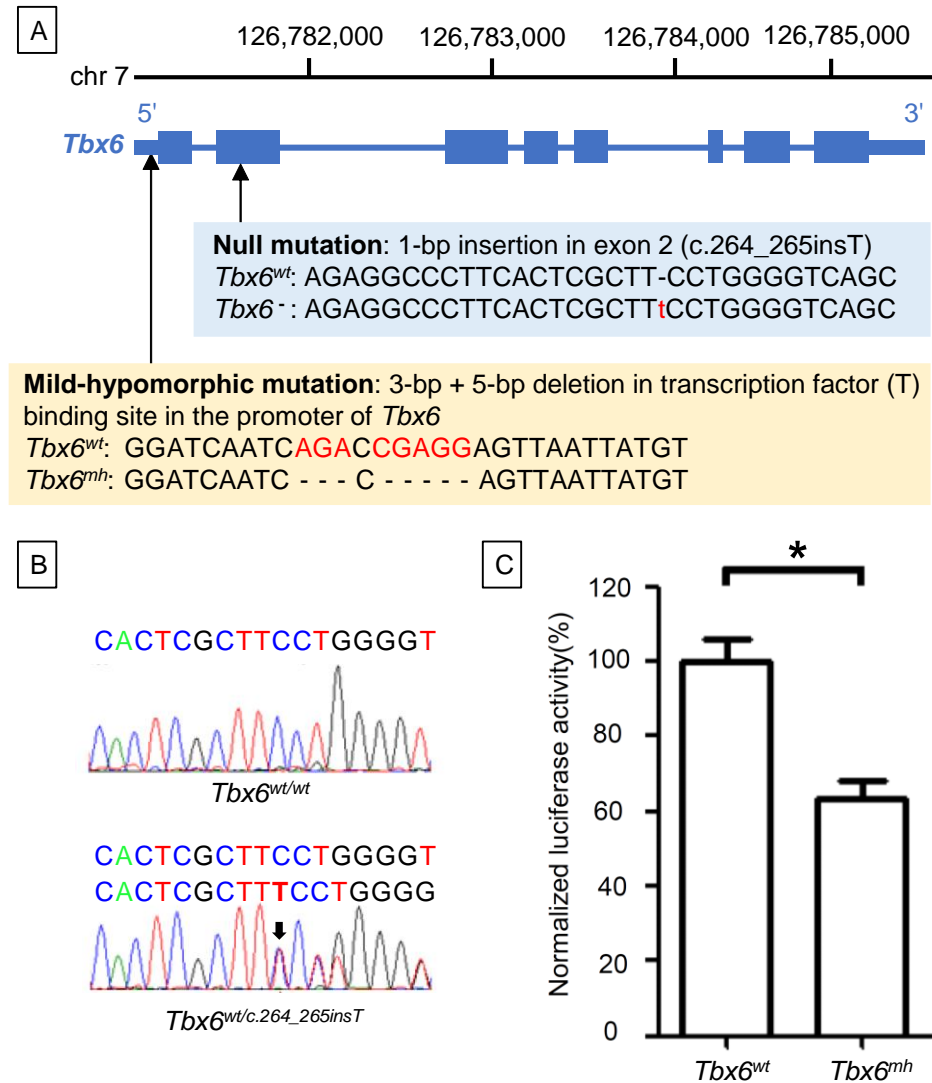


**Figure S6. Distribution of TACS in CS Cohort 1.**

(A) All TACS patients exhibited a consistent phenotype with one or more

hemivertebrae/butterfly vertebrae. The green and white areas indicate the percentages of TACS and non-TACS patients, respectively. The percentages indicate the proportion of TACS patients in CS patients, with particular phenotypes. TACS only existed in CS patients with hemivertebrae or butterfly vertebrae (16.2% [33/204] in CS with hemivertebrae/butterfly vertebrae, 0% [0/127] in CS without hemivertebrae/butterfly vertebrae;  $P=4.9\times 10^{-8}$ ; OR, 1.2; 95% CI, 1.1-1.3), and a relatively high frequency of TACS appeared in patients with a low number of hemivertebrae or butterfly vertebrae ( $P=0.02$  using Pearson  $\chi^2$  test). **(B)** Percentage of each type of rib anomaly and malformation in the TACS and non-TACS groups. The percentages also indicate the proportion of TACS patients in CS patients with particular phenotypes. The frequency of TACS was compared based on the extent of vertebral and rib anomalies in Cohort 1. CS patients with one hemivertebra/butterfly vertebra and simple rib anomaly had the highest proportion of TACS (39.1%, 18 TACS in 46 CS patients), followed by CS patients with one hemivertebra/butterfly vertebra without rib anomaly (17.6%, 6 TACS in 34 CS patients).

Abbreviation: CS, congenital scoliosis; TACS, *TBX6*-associated CS; Non-TACS, non-*TBX6*-associated CS.

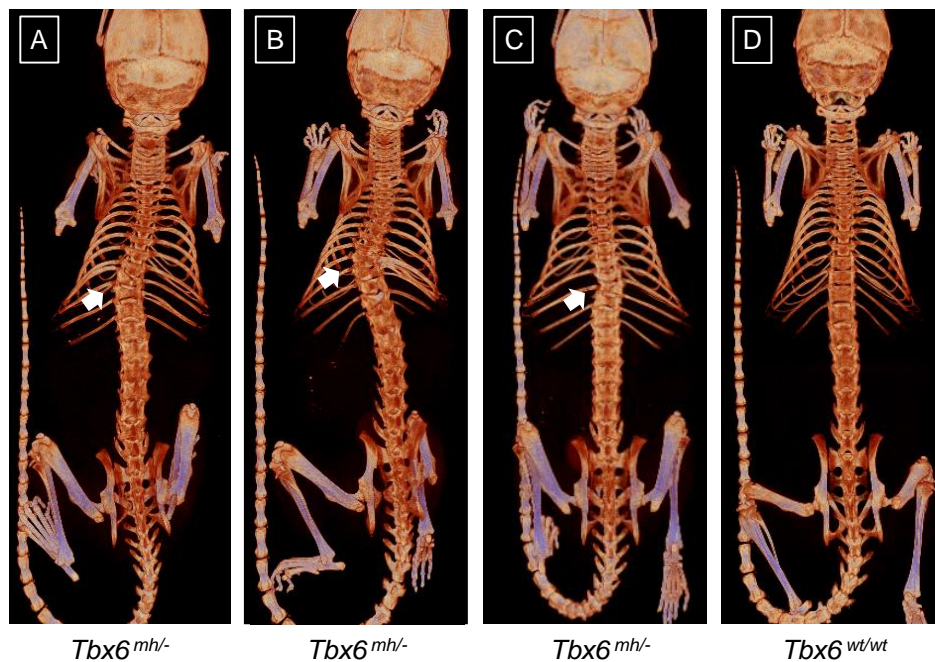


**Figure S7. Mouse *Tbx6* LoF and Mild-hypomorphic Mutation Generated by CRISPR/Cas9.**

(A) Editing of the *Tbx6* gene in the mouse genome (GRCm38/mm10).<sup>7</sup> The *Tbx6* LoF mutation was prepared by introducing a frameshift mutation into exon 2, and the mild-hypomorphic (mh) allele of *Tbx6* was prepared by editing the T binding site in the promoter of *Tbx6*. (B) Sanger sequencing verified the insertion into exon 2 of the *Tbx6* gene. The black arrows indicate the nucleotide substitution positions. The homozygous wild-type (wt) allele is shown as the mouse

genome reference. (C) The hypomorphic effect on gene expression was investigated *in vitro* in P19CL6 cells using luciferase reporter assays.<sup>7</sup> The graph shows the mean values of normalized luciferase activity and the standard errors. Gene expression was down-regulated by the *Tbx6*<sup>mh</sup> mutant to approximately 65% of the wild-type gene, which was close to the 70% dosage level of the human *TBX6* mild-hypomorphic allele.<sup>1</sup> \*,  $P < 0.0001$ .

Abbreviation: wt, wild-type; mh, mild-hypomorphic.

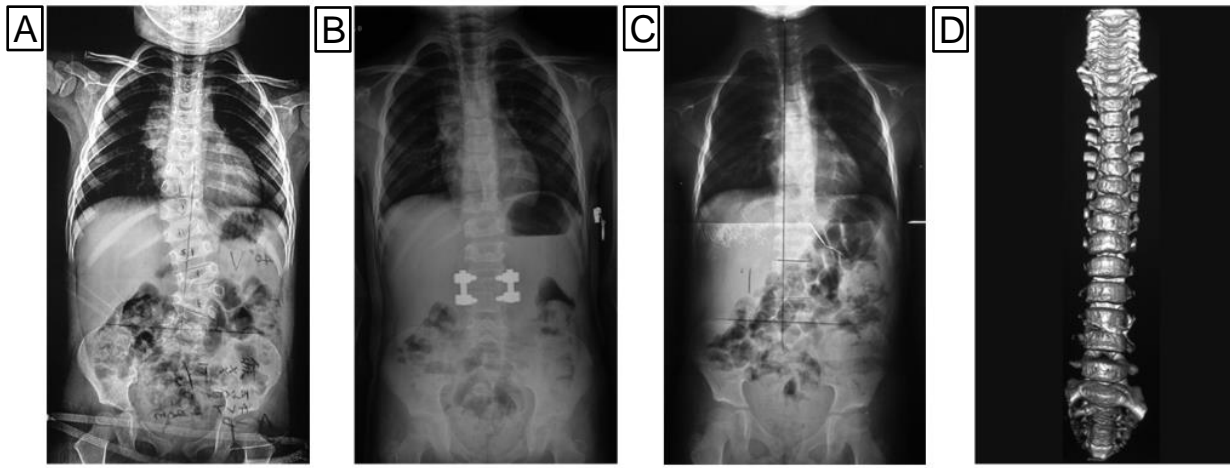


**Figure S8. Spinal Phenotype of *Tbx6* Gene-Edited Mice.**

(A-C) Fifty-two *Tbx6*<sup>mh/-</sup> mice underwent micro-CT scans to obtain 3D skeleton images of the whole body, especially the vertebrae and ribs. Vertebral malformations were involved in 92.3% (48/52) of the *Tbx6*<sup>mh/-</sup> mice, and all vertebral malformations presented as defect of proper

formation in the lower part of the spine, indicated by the white arrows. Additionally, 98.1% (51/52) of the *Tbx6*<sup>mh/-</sup> mice also exhibited a consistent phenotype of rib fusion. **(D)** The wild-type mouse.

Abbreviation: wt, wild-type; mh, mild-hypomorphic.



**Figure S9. Vertebral Malformation in a TACS Patient Treated by Surgical Correction at an Early Age.**

**(A)** Patient XH101 was diagnosed as TACS with a L2 hemivertebra at 3 years of age. Molecular testing indicated that she had a *TBX6* frameshift mutation (c.1250\_1251insT) with the ‘T-C-A’ risk haplotype in the opposite allele. **(B)** The L2 hemivertebra resection was performed with short segment fixation. **(C)** and **(D)** The X-ray and CT reconstruction show a nearly normal spinal column at 4 years post-operation. The instrument was removed at the 3-year follow-up post-operation when the patient had a balanced spine. The spine exhibited good function and a balanced appearance during our follow-up on Sep 26<sup>th</sup>, 2017 (9 years post-operation).

**Table S1. Genetic Analysis Methods in Each Cohort.**

| Cohort                         | Total Number of Patients | Enrollment Center   | <i>TBX6</i> SNV Detection   | 16p11.2 del Detection   |
|--------------------------------|--------------------------|---|---|---|
| <b>The Discovery Set</b>       |                          |   |   |   |
| Cohort 1                       | 345                      | Peking Union Medical College Hospital (PUMCH) in China                    | The entire <i>TBX6</i> gene and its upstream region were amplified and analyzed by Sanger sequencing.   | A genome-wide analysis of copy-number variants was performed in 20 CS patients and their healthy parents with the use of Agilent 1×1M CGH microarrays. A qPCR analysis was conducted in the remaining 325 patients. Digital droplet PCR was used to confirm the 16p11.2 del in XH330 and XH623. Customized CGH microarrays were used to validate/confirm the remaining deletion CNV candidates detected by the qPCR analysis. |
| <b>The Replication Cohorts</b> |                          |   |   |   |
| Cohort 2                       | 142                      | RIKEN and participating hospitals in Japan                                | Sanger sequencing of the entire <i>TBX6</i> gene and its upstream region for 121 patients.<br><i>TBX6</i> variants were also detected by exome sequencing and verified by Sanger sequencing in the remaining 21 patients. | qPCR analysis was conducted in all the patients. CGH microarrays were used to validate/confirm the deletion CNV candidates.   |
| Cohort 3                       | 10                       | Baylor College of Medicine, Drexel University College of Medicine, Boston | <i>TBX6</i> variants were detected by exome sequencing and verified by Sanger sequencing in BH8084.   | Clinical chromosomal microarray analysis (CMA) was conducted for all the patients in Cohort 3   |

---

Children's Hospital,  
Washington,  
University School of  
Medicine, and  
Children's Hospital  
Central California.

---

Sanger sequencing of the  
entire *TBX6* gene and its  
upstream region for the rest  
of the patients.

Abbreviation: SNV, single nucleotide variant; del, deletion; qPCR, quantitative polymerase chain reaction; CGH, comparative genomic hybridization; CNV, copy number variant; CMA, chromosomal microarray analysis.

**Table S2. Genotype and Demographic Information of all TACS Patients in Cohort 1 and Cohort 2.**

|                                   | Patient No. | Gender | Age (Yr) <sup>a</sup> | LoF Mutation at First Allele | Risk Haplotype at Second Allele <sup>b</sup> | Total Number |
|-----------------------------------|-------------|--------|-----------------------|------------------------------|--|--------------|
| <b>Cohort 1</b>                   |             |        |                       |                              |  |              |
| Previously described <sup>c</sup> | XH004       | F      | 4                     | 16p11.2 del                  | T-C-A  | 17           |
|                                   | XH025       | M      | 3                     | 16p11.2 del                  | T-C-A  |              |
|                                   | XH042       | M      | 6                     | 16p11.2 del                  | T-C-A  |              |
|                                   | XH141       | F      | 18                    | 16p11.2 del                  | T-C-A  |              |
|                                   | XH149       | F      | 7                     | 16p11.2 del                  | T-C-A  |              |
|                                   | XH186       | M      | 4                     | 16p11.2 del                  | T-C-A  |              |
|                                   | XH237       | M      | 13                    | 16p11.2 del                  | T-C-A  |              |
|                                   | XH265       | M      | 4                     | 16p11.2 del                  | T-C-A  |              |
|                                   | XH270       | M      | 14                    | 16p11.2 del                  | T-C-A  |              |
|                                   | XH292       | F      | 11                    | 16p11.2 del                  | T-C-A  |              |
|                                   | XH300       | M      | 3                     | 16p11.2 del                  | T-C-A  |              |
|                                   | XH303       | M      | 5                     | 16p11.2 del                  | T-C-A  |              |
|                                   | XR345       | F      | 13                    | 16p11.2 del                  | T-C-A  |              |
|                                   | XR353       | M      | 11                    | 16p11.2 del                  | T-C-A  |              |
|                                   | XR402       | M      | 11                    | 16p11.2 del                  | T-C-A  |              |
|                                   | XR434       | F      | 16                    | 16p11.2 del                  | T-C-A  |              |
|                                   | XR439       | F      | 7                     | 16p11.2 del                  | T-C-A  |              |
|                                   | XH101       | F      | 3                     | c.1250_1251insT              | T-C-A  | 5            |
|                                   | XH122       | M      | 5                     | c.266_267insC                | T-C-A  |              |
|                                   | XH170       | M      | 5                     | c.704_705insG                | T-C-A  |              |
|                                   | XH286       | M      | 2                     | c.1169_1170insC              | T-C-A  |              |
|                                   | XR341       | M      | 2                     | c.1179_1180delAG             | T-C-A  |              |
|                                   | XH148       | F      | 16                    | c.844C>T<br>(p.R282*)        | T-C-A  | 1            |

|                                   | Patient No. | Gender | Age (Yr) <sup>a</sup> | LoF Mutation at First Allele    | Haplotype at Opposite Allele <sup>b</sup> | Total Number |
|-----------------------------------|-------------|--------|-----------------------|---------------------------------|---|--------------|
| <b>Cohort 1 (continued)</b>       |             |        |                       |                                 |   |              |
| Newly recruited                   | XH139       | M      | 3                     | 16p11.2 del                     | T-C-A                                     | 9            |
|                                   | XH330       | M      | 4                     | 16p11.2 del                     | T-C-A                                     |              |
|                                   | XH468       | M      | 3                     | 16p11.2 del                     | T-C-A                                     |              |
|                                   | XH480       | F      | 11                    | 16p11.2 del                     | T-C-A                                     |              |
|                                   | XH522       | M      | 8                     | 16p11.2 del                     | T-C-A                                     |              |
|                                   | XH529       | M      | 14                    | 16p11.2 del                     | T-C-A                                     |              |
|                                   | XH605       | F      | 11                    | 16p11.2 del                     | T-C-A                                     |              |
|                                   | XH623       | M      | 1                     | 16p11.2 del                     | T-C-A                                     |              |
|                                   | XH636       | F      | 9                     | 16p11.2 del                     | T-C-A                                     |              |
|                                   | XH625       | M      | 7                     | c.933C>A (p.C311*)              | T-C-A                                     | 1            |
| <b>Cohort 2</b>                   |             |        |                       |                                 |   |              |
| Previously described <sup>d</sup> | CS029       | F      | 5                     | 16p11.2 del                     | T-C-A                                     | 5            |
|                                   | A1031       | M      | 13                    | 16p11.2 del                     | T-C-A                                     |              |
|                                   | CS119       | M      | 9                     | 16p11.2 del                     | T-C-A                                     |              |
|                                   | A6011       | M      | 13                    | 16p11.2 del                     | T-C-A                                     |              |
|                                   | A1025       | F      | 6                     | 16p11.2 del                     | T-C-A                                     |              |
|                                   | A1008       | M      | 12                    | c.156delG                       | T-C-A                                     | 2            |
|                                   | CS121       | F      | 17                    | c.935_936insGA                  | T-C-A                                     |              |
|                                   | CS024       | F      | 17                    | c.699G>A (p.W233*)              | T-C-A                                     | 1            |
|                                   | CS131       | F      | 6                     | c.333G>T (p.M111I) <sup>e</sup> | T-C-A                                     | 1            |
| Newly recruited                   | A1042       | F      | 15                    | 16p11.2del                      | T-C-A                                     | 5            |
|                                   | A1076       | F      | 9                     | 16p11.2del                      | T-C-A                                     |              |
|                                   | A9022       | M      | 7                     | 16p11.2del                      | T-C-A                                     |              |
|                                   | S1275       | M      | 14                    | 16p11.2del                      | T-C-A                                     |              |
|                                   | S1325       | F      | 13                    | 16p11.2del                      | T-C-A                                     |              |
|                                   | A1107       | F      | 15                    | c.119-1G>A                      | T-C-A                                     | 1            |

Abbreviation: TACS, *TBX6*-associated congenital scoliosis; No., number; yr, year(s); LoF, Loss-

of-Function; del, deletion; ins, insertion.

<sup>a</sup> Age at the time of enrollment.

<sup>b</sup> The haplotype defined by three common *TBX6* SNPs (wild-type/mutant): rs2289292 (C/T) - rs3809624 (T/C) - rs3809627 (C/A).

<sup>c</sup> Patients described in our previous study.<sup>1</sup>

<sup>d</sup> Patients described previously.<sup>2</sup>

<sup>e</sup> The novel missense mutation of c.333G>T (p.M111I) in the Japanese cohort was confirmed as a loss-of-function variant *in vitro*.<sup>2</sup>

**Table S3. Mouse Phenotype with Different *Tbx6* Genotypes.**

| <b>Genotype</b>   | <b><i>Tbx6</i><sup>wt/wt</sup></b> | <b><i>Tbx6</i><sup>wt/-</sup></b> | <b><i>Tbx6</i><sup>wt/mh</sup></b> | <b><i>Tbx6</i><sup>mh/mh</sup></b> | <b><i>Tbx6</i><sup>mh/-</sup></b> | <b><i>P</i> value</b> | <b>OR</b> | <b>95% CI</b> |
|---|------------------------------------|-----------------------------------|------------------------------------|------------------------------------|-----------------------------------|-----------------------|-----------|---------------|
| Total number, No. (%)   | 10                                 | 10                                | 10                                 | 10                                 | 52                                |                       |           |               |
| With vertebral malformation, No. (%)                            | 0 (0)                              | 0 (0)                             | 0 (0)                              | 0 (0)                              | 48 (92.3)                         | 9.3×10 <sup>-9</sup>  | 13.0      | 5.1-33.3      |
| With vertebral malformation in lower part of the spine, No. (%) | 0 (0)                              | 0 (0)                             | 0 (0)                              | 0 (0)                              | 48 (92.3)                         | 9.3×10 <sup>-9</sup>  | 13.0      | 5.1-33.3      |
| With defect of vertebral formation, No. (%)                     | 0 (0)                              | 0 (0)                             | 0 (0)                              | 0 (0)                              | 36 (69.2)                         | 4.9×10 <sup>-5</sup>  | 3.3       | 2.2-4.9       |
| With rib fusion, No. (%)  | 0 (0)                              | 0 (0)                             | 0 (0)                              | 0 (0)                              | 51 (98.1)                         | 1.0×10 <sup>-10</sup> | 52.0      | 7.5-362.2     |

Abbreviation: wt, wild-type; mh, mild-hypomorphic; OR, Odds ratio; CI, confidence interval; No., number.

Italic *P* value<0.05 is statistically significant.

**Table S4. Parameters in the Multivariable TACS Model.**

| Variable                                     | Regression coefficient | <i>P</i> value       | OR   | 95% CI   |          |
|--|------------------------|----------------------|------|----------|----------|
| Segmented hemi-/butterfly vertebrae in T8-S5 | 2.799                  | 2.5×10 <sup>-5</sup> | 16.4 | 4.5-60.4 |          |
| Number of vertebral malformations            | -0.292                 | 0.08                 | 0.8  | 0.5-1.0  |          |
| Intraspinal defects                          | -2.537                 | 0.02                 | 0.08 | 0.01-0.6 |          |
| Rib defect                                   | Without                | -1.376               | 0.01 | 0.3      | 0.1-0.7  |
|  | Complex                | -2.224               | 0.04 | 0.1      | 0.01-0.9 |
|  | Simple                 |                      | 0.01 |          |          |
| Intercept                                    | -1.880                 | 0.02                 | 0.2  |          |          |

Abbreviations: OR, odds ratio; CI, confidence interval; hemi, hemivertebra; T, thoracic vertebra; L, lumbar vertebra; S, sacral vertebra.

*Italic P* value<0.05 is statistically significant.

**Table S5. Using Points≥3 as the Cutoff Value in the Discovery Cohort 1.**

|       |          | Expected |          |       |
|-------|----------|----------|----------|-------|
|       |          | TACS     | Non-TACS | Total |
| Exact | TACS     | 31       | 2        | 33    |
|       | Non-TACS | 27       | 271      | 298   |
|       | Total    | 57       | 274      | 331   |

Abbreviation: CS, congenital scoliosis; TACS, *TBX6*-associated CS; Non-TACS, non-*TBX6*-associated CS.

Sensitivity (SE): 93.9%. Specificity (SP): 90.9%. Accuracy: 91.2%.

**Table S6. Using Points $\geq$ 3 as the Cutoff Value in the Validation Cohort 2.**

|       |          | Expected |          |       |
|-------|----------|----------|----------|-------|
|       |          | TACS     | Non-TACS | Total |
| Exact | TACS     | 12       | 3        | 15    |
|       | Non-TACS | 40       | 87       | 127   |
|       | Total    | 52       | 90       | 142   |

Abbreviation: CS, congenital scoliosis; TACS, *TBX6*-associated CS; Non-TACS, non-*TBX6*-associated CS.

Sensitivity (SE): 80.0%. Specificity (SP): 68.5%. Accuracy: 69.7%.

## References in Supplementary Materials

1. Wu N, Ming X, Xiao J, et al. *TBX6* null variants and a common hypomorphic allele in congenital scoliosis. *N Engl J Med*. 2015;372(4):341-350.
2. Takeda K, Kou I, Kawakami N, et al. Compound Heterozygosity for Null Mutations and a Common Hypomorphic Risk Haplotype in *Tbx6* Causes Congenital Scoliosis. *Human Mutation*. 2017.
3. McMaster MJ, Ohtsuka K. The natural history of congenital scoliosis. A study of two hundred and fifty-one patients. *J Bone Joint Surg Am*. 1982;64(8):1128-1147.
4. Marks DS, Qaimkhani SA. The natural history of congenital scoliosis and kyphosis. *Spine (Phila Pa 1976)*. 2009;34(17):1751-1755.
5. McMaster MJ, David CV. Hemivertebra as a cause of scoliosis. A study of 104 patients. *J Bone Joint Surg Br*. 1986;68(4):588-595.
6. Tsirikos AI, McMaster MJ. Congenital anomalies of the ribs and chest wall associated with congenital deformities of the spine. *J Bone Joint Surg Am*. 2005;87(11):2523-2536.
7. Yang N, Wu N, Zhang L, et al. *TBX6* compound inheritance leads to congenital vertebral malformations in humans and mice. *Human Molecular Genetics*. In press.
8. Li L, Gong H, Yu H, et al. Knockdown of nucleosome assembly protein 1-like 1 promotes dimethyl sulfoxide-induced differentiation of P19CL6 cells into cardiomyocytes. *J Cell Biochem*. 2012;113(12):3788-3796.
9. Sullivan LM, Massaro JM, D'Agostino RB, Sr. Presentation of multivariate data for clinical use: The Framingham Study risk score functions. *Stat Med*. 2004;23(10):1631-1660.
10. Reiser B. Measuring the effectiveness of diagnostic markers in the presence of

measurement error through the use of ROC curves. *Stat Med.* 2000;19(16):2115-2129.